

1 CLAIMS

2
3 1. One or more computer readable media having stored thereon a
4 plurality of instructions that, when executed by one or more processors, causes the
5 one or more processors to perform acts including:

6 receiving the audio portion of a sporting event;
7 classifying a set of segments of the audio portion as excited speech;
8 classifying a set of frame groupings as including baseball hits;
9 combining the set of segments and the set of frame groupings to identify
10 probabilities for each segment that the segment is an exciting segment; and
11 saving an indication of the set of segments and the corresponding
12 probabilities as meta data corresponding to the sporting event.

13
14 2. One or more computer readable media as recited in claim 1, wherein
15 each segment includes at least ten 0.5-second windows of the audio portion.

16
17 3. One or more computer readable media as recited in claim 1, wherein
18 each frame grouping is a grouping of 25 10-millisecond frames of the audio
19 portion.

20
21 4. One or more computer readable media as recited in claim 1, wherein
22 the classifying a set of segments as excited speech comprises:

23 extracting a first set of features from the audio portion;
24 identifying, based on the first set of features, a plurality of windows of the
25 audio portion that include speech;

1 extracting a second set of features from the audio portion; and
2 identifying, based on the second set of features, which of the plurality of
3 windows include excited speech.
4

5 5. One or more computer readable media as recited in claim 4, wherein
6 extracting the first set of features comprises, for each of a plurality of windows:

7 identifying an average waveform amplitude of the audio portion in a first
8 frequency band of the window;

9 identifying an average waveform amplitude of the audio portion in a second
10 frequency band of the window;

11 concatenating the identified average waveform amplitudes to generate an
12 energy feature of the first set of features; and

13 determining, for an MFCC feature of the first set of features, the Mel-
14 frequency Cepstral coefficient of the window.
15

16 6. One or more computer readable media as recited in claim 5, wherein
17 the identifying a plurality of windows of the audio portion that include speech
18 comprises determining that a window includes speech if the energy feature
19 exceeds a first threshold and the MFCC feature exceeds a second threshold.
20

21 7. One or more computer readable media as recited in claim 4, wherein
22 extracting the second set of features comprises, for each of a plurality of windows:

23 identifying, for each of a plurality of frames in the window, an average
24 waveform amplitude of the audio portion in a first frequency band;
25

1 identifying, for each of the plurality of frames in the window, an average
2 waveform amplitude of the audio portion in a second frequency band;
3 concatenating the identified average waveform amplitudes to generate an
4 energy feature;
5 extracting, as a pitch feature of each frame, the pitch of each frame;
6 identifying a plurality of statistics regarding each window based on the
7 energy features and pitch features of the plurality of frames.

8
9 8. One or more computer readable media as recited in claim 7, wherein
10 the identifying a plurality of statistics further comprises:

11 identifying a maximum energy;
12 identifying an average energy;
13 identifying an energy dynamic range;
14 identifying a maximum pitch;
15 identifying an average pitch; and
16 identifying a dynamic pitch range.

17
18 9. One or more computer readable media as recited in claim 7, wherein
19 the identifying which of the plurality of windows include excited speech
20 comprises identifying a *posterior* probability that the window corresponds to an
21 excited speech class and identifying a *posterior* probability that the window
22 corresponds to a non-excited speech class, and classifying the window in the class
23 with the higher *posterior* probability.
24
25

1 **10.** One or more computer readable media as recited in claim 4, further
2 comprising instructions that cause the one or more processors to perform acts
3 including outputting an excited speech probability for each of the segments that
4 include excited speech, the excited speech probability for a segment indicating a
5 likelihood that the segment includes excited speech.

6
7 **11.** One or more computer readable media as recited in claim 1, wherein
8 the classifying a set of frame groupings as including baseball hits:

9 extracting, for each frame in a multiple-frame grouping, a set of features
10 from the audio portion;

11 comparing the set of features from the multiple-frame groupings to a set of
12 templates; and

13 identifying, for each of the multiple-frame groupings, a probability that the
14 grouping includes a baseball hit based on how well the grouping matches the set of
15 templates.

16
17 **12.** One or more computer readable media as recited in claim 11,
18 wherein the extracting comprises, for each frame:

19 identifying an average waveform amplitude of the audio portion in a first
20 frequency band of the frame;

21 identifying an average waveform amplitude of the audio portion in a second
22 frequency band of the frame;

23 concatenating the identified average waveform amplitudes to generate a
24 first energy feature of the set of features;

identifying an average waveform amplitude of the audio portion in a third frequency band of the frame; and

using, as a second energy feature of the set of features, the average waveform amplitude of the third frequency band.

13. One or more computer readable media as recited in claim 12, further comprising instructions that cause the one or more processors to perform acts including:

generating a third energy feature for each frame by normalizing the first energy feature based on the first energy feature of the eighth frame of the multiple-frame grouping; and

generating a fourth energy feature for each frame by normalizing the second energy feature based on the second energy feature of the eighth frame of the multiple-frame grouping.

14. One or more computer readable media as recited in claim 1, wherein the combining comprises generating, for each of the set of segments, a weighted sum of a probability that the segment includes excited speech and a probability that a frame grouping within the segment includes a baseball hit.

15. One or more computer readable media as recited in claim 1, wherein the combining comprises adjusting a probability that a segment of the set of segments includes excited speech based on a probability that the segment includes a baseball hit.

1 **16.** A method comprising:
2 receiving a program including both audio and video;
3 receiving meta data corresponding to the program; and
4 rendering, based on the meta data, portions of the program as a summary of
5 the program.

6
7 **17.** A method as recited in claim 16, wherein the rendering comprises
8 displaying the video of the portions and playing the audio of the portions.

9
10 **18.** A method as recited in claim 16, wherein the meta data comprises a
11 probability indicator, for each of a plurality of portions of the program, that
12 identifies a probability that the corresponding portion is an exciting portion of the
13 program.

14
15 **19.** A method as recited in claim 18, wherein the rendering comprises
16 selecting the plurality of portions that have probability indicators that exceed a
17 threshold value, and rendering the selected portions as the summary.

18
19 **20.** A method as recited in claim 16, wherein the receiving a program
20 and the receiving meta data comprise receiving both the program and the meta
21 data from the same source.

22
23 **21.** A method as recited in claim 16, wherein the receiving meta data
24 comprises receiving the meta data from a remote source via a network.
25

1 **22.** A method as recited in claim 21, wherein the network comprises the
2 Internet.

3
4 **23.** A method as recited in claim 16, wherein the receiving meta data
5 comprises receiving meta data generated manually.

6
7 **24.** A method as recited in claim 16, wherein the receiving meta data
8 comprises receiving meta data generated automatically.

9
10 **25.** A method as recited in claim 16, wherein the meta data comprises a
11 plurality of probabilities, each corresponding to a segment of the program, the
12 probabilities representing a probabilistic combination of sports-specific events and
13 sports-generic events identified in the program.

14
15 **26.** A method as recited in claim 25, wherein the sports-specific events
16 comprise baseball hits, and wherein the sports-generic events comprise excited
17 speech.

18
19 **27.** One or more computer readable media including a computer
20 program that is executable by a processor to perform the method recited in claim
21 16.

22 **28.** A system comprising:
23
24 a content provider to make programming content available to requesting
25 clients;

a meta data provider to make meta data, corresponding to the programming content, available to the requesting clients, wherein the meta data identifies, for each of a plurality of portions of the programming content, an indicator of a likelihood that the corresponding portion is an exciting portion of the programming content; and

a plurality of receivers coupled to receive the programming content from the content provider and the meta data from the meta data provider.

29. A system as recited in claim 28, wherein the content provider and the meta data provider are the same.

30. A system as recited in claim 28, wherein the content provider and the meta data provider are coupled to the plurality of receivers via the Internet.

31. A system as recited in claim 28, wherein the plurality of receivers are further to render, based on the meta data, portions of the programming content as a summary of the programming content.

32. A system as recited in claim 28, wherein the meta data comprises a probability indicator, for each of a plurality of portions of the programming content, that identifies a probability that the corresponding portion is an exciting portion of the program.

1 **33.** A system as recited in claim 32, wherein the plurality of receivers
2 are further to select, from the plurality of portions, portions that have probability
3 indicators that exceed a threshold value, and render the selected portions as the
4 summary.

5
6 **34.** A system as recited in claim 28, wherein the meta data is generated
7 manually.

8
9 **35.** A system as recited in claim 28, wherein the meta data is generated
10 automatically.

11
12 **36.** A system as recited in claim 28, wherein the meta data comprises a
13 plurality of probabilities, each corresponding to a segment of the programming
14 content, the probabilities representing a probabilistic combination of sports-
15 specific events and sports-generic events identified in the programming content.

16
17 **37.** A system as recited in claim 36, wherein the sports-specific events
18 comprise baseball hits, and wherein the sports-generic events comprise excited
19 speech.

20
21 **38.** A method of automatically summarizing a program, the method
22 comprising:

23 identifying a plurality of sports-generic events from the audio of the
24 program;
25

1 identifying a plurality of sports-specific events from the audio of the
2 program; and

3 identifying, by combining the sports-generic events and the sports-specific
4 events, a set of portions of the program as a summary of the program.

5
6 **39.** A method as recited in claim 38, further comprising transmitting the
7 set of portions to a client computer as the summary of the program.

8
9 **40.** A method as recited in claim 39, wherein the transmitting comprises
10 transmitting the set of portions via the Internet.

11
12 **41.** A method as recited in claim 38, wherein the sports-specific events
13 comprise baseball hits, and wherein the sports-generic events comprise excited
14 speech.

15
16 **42.** A method as recited in claim 38, wherein the program includes both
17 an audio portion and a video portion.

18
19 **43.** One or more computer readable media including a computer
20 program that is executable by a processor to perform the method recited in claim
21 38.

22
23 **44.** A method comprising:
24 analyzing audio data of a program to identify a first plurality of portions of
25 the program each including excited speech;

1 analyzing the audio data to identify a second plurality of portions of the
2 program each including a potential baseball hit; and

3 combining the first plurality of portions and the second plurality of portions
4 to identify a set of segments of the program and a likelihood, for each of the
5 segments in the set, that the segment is an exciting part of the program.

6
7 **45.** A method as recited in claim 44, wherein the analyzing audio data to
8 identify the first plurality of portions comprises:

9 extracting a first set of features from the audio data;

10 identifying, based on the first set of features, a plurality of windows of the
11 audio data that include speech;

12 extracting a second set of features from the audio data; and

13 identifying, based on the second set of features, which of the plurality of
14 windows include excited speech.

15
16 **46.** A method as recited in claim 45, wherein extracting the first set of
17 features comprises, for each of a plurality of windows:

18 identifying an average waveform amplitude of the audio data in a first
19 frequency band of the window;

20 identifying an average waveform amplitude of the audio data in a second
21 frequency band of the window;

22 concatenating the identified average waveform amplitudes to generate an
23 energy feature of the first set of features; and

24 determining, for an MFCC feature of the first set of features, the Mel-
25 frequency Cepstral coefficient of the window.

1
2 **47.** A method as recited in claim 46, wherein each window comprises
3 0.5 seconds.

4
5 **48.** A method as recited in claim 46, wherein the identifying a plurality
6 of windows of the audio data that include speech comprises determining that a
7 window includes speech if the energy feature exceeds a first threshold and the
8 MFCC feature exceeds a second threshold.

9
10 **49.** A method as recited in claim 45, wherein extracting the second set
11 of features comprises, for each of a plurality of windows:

12 identifying, for each of a plurality of frames in the window, an average
13 waveform amplitude of the audio data in a first frequency band;

14 identifying, for each of the plurality of frames in the window, an average
15 waveform amplitude of the audio data in a second frequency band;

16 concatenating the identified average waveform amplitudes to generate an
17 energy feature;

18 extracting, as a pitch feature of each frame, the pitch of each frame; and

19 identifying a plurality of statistics regarding each window based on the
20 energy features and pitch features of the plurality of frames.

21
22 **50.** A method as recited in claim 49, wherein the identifying a plurality
23 of statistics further comprises:

24 identifying a maximum energy;

25 identifying an average energy;

1 identifying an energy dynamic range;
2 identifying a maximum pitch;
3 identifying an average pitch; and
4 identifying a dynamic pitch range.
5

6 **51.** A method as recited in claim 49, wherein the identifying which of
7 the plurality of windows include excited speech comprises identifying a *posterior*
8 probability that the window corresponds to an excited speech class and identifying
9 a *posterior* probability that the window corresponds to a non-excited speech class,
10 and classifying the window in the class with the higher *posterior* probability.
11

12 **52.** A method as recited in claim 51, wherein the *posteriori* probabilities
13 are identified using a parametric machine.
14

15 **53.** A method as recited in claim 51, wherein the *posteriori* probabilities
16 are identified using a non-parametric machine.
17

18 **54.** A method as recited in claim 51, wherein the *posteriori* probabilities
19 are identified using a semi-parametric machine.
20
21
22
23
24
25

1 identifying an average waveform amplitude of the audio data in a second
2 frequency band of the frame;

3 concatenating the identified average waveform amplitudes to generate a
4 first energy feature of the set of features;

5 identifying an average waveform amplitude of the audio data in a third
6 frequency band of the frame; and

7 using, as a second energy feature of the set of features, the average
8 waveform amplitude of the third frequency band.

9
10 **59.** A method as recited in claim 58, further comprising:

11 generating a third energy feature for each frame by normalizing the first
12 energy feature based on the first energy feature of the eighth frame of the multiple-
13 frame grouping; and

14 generating a fourth energy feature for each frame by normalizing the
15 second energy feature based on the second energy feature of the eighth frame of
16 the multiple-frame grouping.

17
18 **60.** A method as recited in claim 44, wherein the combining comprises
19 generating, for each of the first plurality of portions, a weighted sum of a
20 probability that the portion includes excited speech and a probability that the
21 portion includes a baseball hit.

1 **61.** A method as recited in claim 44, wherein the combining comprises
2 adjusting the probability that a portion includes excited speech based on the
3 probability that the portion includes a baseball hit.

4
5 **62.** A system comprising:
6 a feature extractor to extract a plurality of audio features from programming
7 content;
8 an excited speech classification subsystem to identify, based on a sub-set of
9 the audio features, a set of segments of the programming content and
10 corresponding probabilities that the segments include excited speech;
11 a baseball hit detection subsystem to identify, based on another sub-set of
12 the audio features, a set of frame groupings of the programming content and
13 corresponding probabilities that the frame groupings include baseball hits; and
14 a probabilistic fusion subsystem to combine the probabilities that the
15 segments include excited speech and the probabilities that the frame groupings
16 include baseball hits, and to generate a probability that portions of the
17 programming content are exciting based on the combination.

18
19 **63.** A system as recited in claim 62, wherein the excited speech
20 classification subsystem is to identify the set of segments by:
21 identifying, based on a first set of the plurality of audio features, a plurality
22 of windows of the programming content that include speech;
23 identifying, based on a second set of the plurality of audio features, which
24 of the plurality of windows include excited speech.

25

1 **64.** A system as recited in claim 62, wherein the baseball hit detection
2 subsystem is to identify the set of frame groupings by:

3 combining, for each of a plurality of multiple-frame groupings, a set of
4 features from the programming content;

5 comparing the sets of features from the multiple-frame groupings to a set of
6 templates; and

7 identifying, for each of the multiple-frame groupings, a probability that the
8 grouping includes a baseball hit based on how well the grouping matches the set of
9 templates.

10
11 **65.** A system comprising:

12 a receiving device to receive a sporting event;

13 a user interface to receive, from a user, an indication of a desired summary
14 level for the sporting event; and

15 a processing subsystem to identify which portions of the sporting event to
16 present to the user based at least in part on both the desired summary level and
17 meta data corresponding to the sporting event, the meta data identifying a
18 likelihood of each of a plurality of portions of the sporting event being exciting
19 based at least in part on the presence of both excited speech and ball hits within
20 the sporting event.